A Novel Susceptibility Locus for Type 1 Diabetes on Chr12q13 Identified by a Genome-Wide Association Study

Hakon Hakonarson,1,2 Hui-Qi Qu,3 Jonathan P. Bradfield,1 Luc Marchand,3 Cecilia E. Kim,1 Joseph T. Glessner,1 Rosemarie Grabs,3 Tracy Casalunovo,1 Shayne P. Taback,4 Edward C. Frackelton,1 Andrew W. Eckert,1 Kiran Annaiah,1 Margaret L. Lawson,5 F. George Ottieno,1 Erin Santa,1 Julie L. Shaner,1 Ryan M. Smith,1 Chioma C. Onyiah,1 Robert Skraban,1 Rosetta M. Chiavacci,1 Luke J. Robinson,1 Charles A. Stanley,6 Susan E. Kirsch,7 Marcella Devoto,2,8 Dimitri S. Monos,9,10 Struan F.A. Grant,1,2 and Constantin Polychronakos3

OBJECTIVE—In stage 1 of our genome-wide association (GWA) study for type 1 diabetes, one locus at 16p13 was detected at a genome-wide significance threshold (P = 1.03 × 10\(^{-18}\)) and confirmed in two additional cohorts. Here we describe the results of testing, in these additional cohorts, 23 loci that were next in rank of statistical significance.

RESEARCH DESIGN AND METHODS—Two independent cohorts were studied. The Type 1 Diabetes Genetics Consortium replication cohort consisted of 549 families with at least one child diagnosed with diabetes (946 total affected) and DNA from both parents. The Canadian replication cohort consisted of 364 nuclear family trios with one type 1 diabetes–affected offspring and two parents (1,092 individuals).

RESULTS—One locus at 12q13, with the highest statistical significance among the 23, was confirmed. It involves type 1 diabetes association with the minor allele of rs1701704 (P = 9.13 × 10\(^{-12}\), OR 1.25 [95% CI 1.12–1.40]).

CONCLUSIONS—We have discovered a type 1 diabetes locus at 12q13 that is replicated in an independent cohort of type 1 diabetic patients and confers a type 1 diabetes risk comparable with that of the 16p13 locus we recently reported. These two loci are identical to two loci identified by the whole-genome association study of the Wellcome Trust Case-Control Consortium, a parallel independent discovery that adds further support to the validity of the GWA approach. Diabetes 57:1143–1146, 2008

Type 1 diabetes, a multifactorial disease with a strong genetic component, is due to the autoimmune destruction of pancreatic \(\beta\)-cells. The major type 1 diabetes susceptibility locus, mapping to the \(HLA\) class II genes at \(6p21\) (1) and encoding highly polymorphic antigen-presenting proteins, accounts for less than half of genetic type 1 diabetes risk (2). The remainder is accounted for by many loci with more modest effects, of which only a few are known. These include (1) the insulin \((INS)\) VNTR (variable number tandem repeats) (3) modulating thymic expression of and tolerance to insulin, a major type 1 diabetes autoantigen (4,5); (2) the Arg620Trp single nucleotide polymorphism (SNP) at \(PTPN22\), which affects the function of a negative regulator of T-cell receptor signaling (6); (3) noncoding SNPs at \(IL2RA\) (7–9), which encodes the \(\alpha\)-chain of the IL2 receptor complex (CD25), an important modulator of immunity; and (4) variants in the \(CTLA4\) locus (10) whose protein product transmits inhibitory signals to attenuate T-cell activation. It is worth noting that all of these type 1 diabetes–associated genes are expressed in cells with immune function, and all except \(INS\) have been associated with other autoimmune disorders. Together, these loci explain over half of the genetic type 1 diabetes risk, the remaining of which is composed of loci whose number and effect size are still unknown. It is also worth remarking that these known loci represent the “low-hanging fruit,” involving candidate genes whose importance in autoimmunity was already known, thus limiting the value of genetic discovery to generate previously unsuspected physiological insights.

The recent development of high-throughput genotyping array technologies has enabled us (11) and others (12,13) to perform genome-wide association (GWA) studies in search of the remaining type 1 diabetes loci. The first successful use in type 1 diabetes involved the screening of 12,000 nonsynonymous SNPs, which found type 1 diabetes association with rs1990760, involving an Ala946Thr substitution on the \(IFIH1\) (interferon-induced with helicase C domain 1) gene (14). More recently, the Wellcome Trust Case-Control Consortium (WTCCC) tested 2,000 type 1 diabetic case and 3,000 control subjects for 500 k SNPs (Affymetrix GeneChip) (12), and four novel associations...
TABLE 1

The type 1 diabetes–associated 12q13 SNPs in two independent cohorts

<table>
<thead>
<tr>
<th>SNP</th>
<th>Minor allele</th>
<th>Frequency</th>
<th>Informative families*</th>
<th>S†</th>
<th>E(S)‡</th>
<th>Var(S)§</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs773107</td>
<td>G</td>
<td>0.322</td>
<td>217</td>
<td>329</td>
<td>290</td>
<td>208</td>
<td>2.73</td>
<td>0.006400</td>
</tr>
<tr>
<td>rs10876864</td>
<td>G</td>
<td>0.417</td>
<td>226</td>
<td>395</td>
<td>350</td>
<td>229</td>
<td>2.98</td>
<td>0.002885</td>
</tr>
<tr>
<td>rs1701704</td>
<td>C</td>
<td>0.346</td>
<td>216</td>
<td>341</td>
<td>296</td>
<td>211</td>
<td>3.09</td>
<td>0.002019</td>
</tr>
<tr>
<td>Canadian set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs773107</td>
<td>G</td>
<td>0.339</td>
<td>200</td>
<td>180</td>
<td>160</td>
<td>75</td>
<td>2.37</td>
<td>0.017736</td>
</tr>
<tr>
<td>rs10876864</td>
<td>G</td>
<td>0.452</td>
<td>212</td>
<td>210</td>
<td>191</td>
<td>87</td>
<td>2.09</td>
<td>0.036293</td>
</tr>
<tr>
<td>rs1701704</td>
<td>C</td>
<td>0.357</td>
<td>216</td>
<td>189</td>
<td>169</td>
<td>81</td>
<td>2.22</td>
<td>0.026268</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs773107</td>
<td>G</td>
<td>0.329</td>
<td>417</td>
<td>509</td>
<td>449</td>
<td>283</td>
<td>3.56</td>
<td>0.000374</td>
</tr>
<tr>
<td>rs10876864</td>
<td>G</td>
<td>0.432</td>
<td>438</td>
<td>605</td>
<td>540</td>
<td>316</td>
<td>3.64</td>
<td>0.000278</td>
</tr>
<tr>
<td>rs1701704</td>
<td>C</td>
<td>0.350</td>
<td>432</td>
<td>530</td>
<td>465</td>
<td>292</td>
<td>3.80</td>
<td>0.000148</td>
</tr>
</tbody>
</table>

*Number of nuclear families informative for (with a non-zero contribution to) FBAT analysis. †Observed allele number in the affected offspring. §Variance of allele distribution among the affected offspring.

(12q24, 12q13, 16p13, and 18p11) were solidly replicated in 4,000 case and 5,000 control subjects, plus an additional 2,997 type 1 diabetic family trios, by the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory (13). Our stage 1 results, which have recently been published (11), report one novel type 1 diabetes locus at Chr16p that reached genome-wide significance in stage 1 and coincided with one of the novel WTCCC findings. We used the Illumina Hap550 array to type 550 k SNPs, selected to tag most of the haplotype structure of the human genome, per phase 2 of the International HapMap project (http://www.hapmap.org). A set of 483 subjects with type 1 diabetes and both parents of each subject in addition to 563 type 1 diabetic probands and 1,146 control subjects, all of European ancestry, were examined. For the case-control data, population stratification was corrected by principal component analysis in the EIGENSTRAT implementation (15).

Both studies independently mapped this association to a 300-kb linkage disequilibrium (LD) block on 16p encompassing two genes of unknown function, KIAA0350 (now renamed C-type lectin 16A, symbol CLEC16A) and LOC642451.

Pending completion of a full stage 2, in which the 1,536 individuals genotyped in HapMap) were included as accuracy control subjects. The call rate of each of the three 12q13 SNPs was 98.2%, and no Mendelian error was found. The genotypes of each SNP were in Hardy-Weinberg equilibrium in the parents.

Statistics. Type 1 diabetes association was tested by the Family Based Association Test (FBAT) software (http://www.biostat.harvard.edu/~lfbat/ fbat.htm) (16), based on the transmission disequilibrium test method. Considering that most of the Type 1 Diabetes Genetics Consortium families have multiple siblings, the option of the empirical variance was used in the FBAT statistics to permit a robust but unbiased test of genetic association. The age-of-onset effect of the SNPs in type 1 diabetic patients was tested by the Kruskal-Wallis test.

RESULTS AND DISCUSSION

As shown in Table 1, the type 1 diabetes association at the 12q13 locus was replicated in both sets, with similar odds ratios. In a combined analysis of the two GWA cohorts and a pool of the two replication cohorts by Fisher's method, all three SNPs attained P values ranging from $2.48 \times 10^{-8}$ to $9.13 \times 10^{-10}$ (Table 2). These data provide convincing evidence of a novel type 1 diabetes locus. The other loci fast tracked to stage 2 are shown in supplementary Table 1. While this report was in preparation the WTCCC results

TABLE 2

The combined P values of type 1 diabetes association of the three 12q13 SNPs

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>GWA case-control*</th>
<th>GWA families</th>
<th>Stage 2</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs773107</td>
<td>12</td>
<td>54,655,773</td>
<td>$2.89 \times 10^{-5}$</td>
<td>$7.81 \times 10^{-3}$</td>
<td>$3.74 \times 10^{-4}$</td>
<td>$2.48 \times 10^{-8}$</td>
</tr>
<tr>
<td>rs10876864</td>
<td>12</td>
<td>54,687,352</td>
<td>$8.39 \times 10^{-5}$</td>
<td>$2.97 \times 10^{-4}$</td>
<td>$2.78 \times 10^{-4}$</td>
<td>$2.47 \times 10^{-9}$</td>
</tr>
<tr>
<td>rs1701704</td>
<td>12</td>
<td>54,698,754</td>
<td>$9.89 \times 10^{-6}$</td>
<td>$1.62 \times 10^{-3}$</td>
<td>$1.48 \times 10^{-4}$</td>
<td>$9.13 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

Data are P values. *Corrected by the EIGENSTRAT method (ref. 15).
were published, reporting four novel type 1 diabetes associations, two of which coincide with our own two novel findings at 16p (13) and the one reported here, 12q13. We consider this simultaneous independent discovery of noncandidate loci a potent validation of the GWA approach.

The three SNPs are located in a 250-kb block of tight LD on Chr12q and are highly correlated ($r^2 = 0.608–0.855$). Type 1 diabetes risk was conferred by the minor allele in all three. The OR (95% CI) of the haplotype of the three SNPs’ minor alleles was 1.25 (1.12–1.40). The genotypic association of the SNPs best fits an additive model of susceptibility on a log scale (supplementary Table 2). There was a trend, with borderline statistical significance, of an effect on age-of-onset of type 1 diabetes (supplementary Table 3): homozygotes of the risk allele may have the disease onset 1 year earlier than the homozygotes of the protective allele. This is clearly the same locus as the one reported by Todd et al. (13), as the type 1 diabetes-associated rs2292239 is in tight LD with our markers (supplementary Fig. 2). In addition, imputed genotype counts for our three significant markers in the WTCCC data showed highly significant type 1 diabetes association, as shown in a meta-analysis of the two studies for the 12q13 locus (supplementary Table 4). Significance was increased by several orders of magnitude. None of the other loci tested in this report (supplementary Table 1) achieved statistical significance as a result of this joint analysis.

The associated 250-kb LD block encompasses several genes that now become candidates for the type 1 diabetes association (Fig. 1 and supplementary Fig. 3). RAB5B (MIM:179514) encodes a member of the RAS oncogene family that may be involved in vesicular trafficking at the plasma membrane (17). SUOX (MIM:606887) encodes a liver-specific sulfite oxidase involved in the degradation of sulfur-containing amino acids. Neither gene has known function in immunity or pancreatic development. IKZF4 (MIM:606239) encodes a better functional candidate, a zinc finger protein specifically expressed in lymphocytes and implicated in the control of lymphoid development. ERBB3 (MIM:190151) encodes a member of the epidermal

FIG. 1. Type 1 diabetes association in the 12q13 region. The combined results of type 1 diabetes association tests of the case-control cohort and the family cohort is shown. The red bars, corresponding to $-\log_{10}(P)$ values, represent the SNPs of significant type 1 diabetes association with $P < 0.05$. The LD map is based on our genotyping data of the family cohort. D’ values (%) are shown in the boxes, and for the empty boxes D’ = 1. The gray scale represents the $r^2$ values. The type 1 diabetes-associated SNPs map to an LD region in the middle of the figure, while the strongest association is from the region around the RAB5B gene. The $r^2$ values drastically drop beyond the two ends of the block containing the type 1 diabetes-associated SNPs. Relatively high D’ values continue for an additional ~250 kb in the telomeric direction, but beyond the block shown there is no type 1 diabetes-associated SNP and no HapMap SNPs with substantial $r^2$ values to those in the block shown (extended LD diagram in supplementary Fig. 3).
growth factor receptor family of receptor tyrosine kinases involved in the regulation of cell proliferation or differentiation. Interestingly, it interacts with proteins with important immune functions, such as Lyn, Itk, and Fer (18). CDK2 (MIM: 116953) encodes a member of the Ser/Thr protein kinase family whose activity is regulated by its protein phosphorylation; members of this gene family have been recently implicated in GWA studies of type 2 diabetes (19–22). Fine mapping and functional studies will be required to identify the causative variant and generate functional insights from this genetic finding.

A more immediate conclusion from this finding is the support it provides to the robustness of the GWA approach, through identification of the same loci in two independent studies. Further, it is worth noting that our power to include in this fast-track replication a locus of magnitude comparable with the one reported here was only 38%. This means that our full second stage, plus future stages and combined meta-analyses, are likely to reveal additional such loci.

ACKNOWLEDGMENTS
We gratefully acknowledge the use of DNA samples from the Type 1 Diabetes Genetics Consortium, funded by National Institutes of Health Grant U01-DK62418. This work was funded by the Children’s Hospital of Philadelphia, the Juvenile Diabetes Research Foundation International and Genome Canada through the Ontario Genomics Institute. H.Q.Q. is supported by a fellowship from the Canadian Institutes of Health Research.

We thank all the patients and their parents and the healthy control subjects for their participation in the study.

REFERENCES
1. Todd JA, Bell JI, McDevitt HO: HLA-DQ beta gene contributes to suscepti-


2. Risch N: Assessing the role of HLA-linked and unlinked determinants of


5. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678, 2007


